DOI:
10.1109/ICMLA.2016.0148

Document Version
Peer reviewed version

A Prediction Modelling and Pattern Detection Approach for the First-Episode Psychosis Associated to Cannabis Use

Wajdi Alghamdi*, Daniel Stamate*, Katherine Vang*,
Daniel Stahl†, Marco Colizzi‡, Giada Tripoli‡, Diego Quattrone§, Olesya Ajnakina‡, Robin M. Murray ‡and Marta Di Forti §
*Data Science & Soft Computing Lab, and Department of Computing, Goldsmiths, University of London
†Department of Biostatistics, Institute of Psychiatry, Psychology and Neuroscience, King's College London
‡Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London
§MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry,
Psychology and Neuroscience, King's College London

***Abstract*—Over the last two decades, a significant body of research has established a link between cannabis use and psychotic outcomes. In this study, we aim to propose a novel symbiotic machine learning and statistical approach to pattern detection and to developing predictive models for the onset of first-episode psychosis. The data used has been gathered from real cases in cooperation with a medical research institution, and comprises a wide set of variables including demographic, drug-related, as well as several variables specifically related to the cannabis use. Our approach is built upon several machine learning techniques whose predictive models have been optimised in a computationally intensive framework. The ability of these models to predict first-episode psychosis has been extensively tested through large scale Monte Carlo simulations. Our results show that Boosted Classification Trees outperform other models in this context, and have significant predictive ability despite a large number of missing values in the data. Furthermore, we extended our approach by further investigating how different patterns of cannabis use relate to new cases of psychosis, via association analysis and bayesian techniques.**

***Keywords*-Predicting first-episode psychosis, Cannabis use, Precision medicine, Prediction modelling, Classification, Monte Carlo simulation, Association analysis, Bayesian inference**

I. Introduction

A number of US states have already legalised or are in the process of legalising the use of cannabis. Some other countries such as Uruguay had previously done so. Moreover, currently cannabis is one of the most used illicit drugs in the world. However, research established a significant link between cannabis use and psychotic symptoms, and that cannabis use is the most preventable risk factor for psychosis [1]. In this context, any harm caused by the cannabis use, in particular in connection to psychosis, should be quantified.

As such, more recently, researchers sought to understand whether specific patterns of cannabis use (such as potency [2] or age [3]) relate to higher risk of psychotic disorders. One study estimated that nearly a quarter of all new psychosis patients in South London (UK) could be attributed to the use of high-potency, skunk-like cannabis [4]. The same study estimated that the risk of experiencing psychotic disorders is roughly three times higher for those who are daily users of cannabis, compared to those who are not users.

However, there is still scope for further understanding of the links between patterns of cannabis use and psychosis [5]. Most studies so far are limited by incomplete or inconsistent records, or a lack of detailed variables, but also by the methodologies used, which are mainly based on a number of conventional statistical techniques such as hypotheses formulation and verification via statistical tests, logistic regression modelling, etc. These methods are traditionally well recognised and used in medical research, but in many situations they do not match the large potential of the modern machine learning methods.

The field of machine learning has advanced at tremendous pace in recent years, with advanced predictive techniques being developed and improved upon. In order for these technologies to become truly refined, they must be applied for use in a variety of fields and subsequently challenged to find relevant solutions [6]. One such area of application is the field of medical research, which has a wide range of potential uses for machine learning [7]–[9]. Several recent studies have sought to compare a variety of algorithms in predicting patient survival after surgeries such as breast cancer surgery [10]. These studies suggest that machine learning can provide medical research with powerful techniques beyond the traditional statistical approaches mostly used in this area. In biomedical engineering, several recent papers explored the potential for classification algorithms to detect disease. This has led to the publication of additional guidance for medical researchers on how to interpret and question such findings [11]. Last but not least, there is a tremendous interest in current interdisciplinary research into exploiting the power of machine and statistical learning to enable further progress in the new and promising area of Precision Medicine, in which predictive modelling plays a key role in forecasting treatment outcomes, and thus decisively contributes in optimising and personalising treatments for patients [9], [12].

In this study, we propose a novel symbiotic machine learning and statistical approach to pattern detection and to developing predictive models for the onset of first-episode psychosis. The dataset we based our study upon was collected in previously conducted medical studies as described in [4], and comprises a wide set of variables including demographic, drug-related, as well as several other variables with specific information on the participants' history of cannabis use. Prior to the prediction modelling, a significant effort in our work was involved in the data pre-processing due to inherent challenges present in data collected in a case-control study involving many missing values, multiple encodings of related information, a significantly large number of variables, etc.

The prediction modelling phase consisted of investigating several machine learning techniques such as k-Nearest Neighbours, Support Vector Machine with different kernels, Decision Trees, Bagged Trees, Boosted Classification Trees, eXtreme Gradient Boosting and Random Forests, whose predictive models have been optimised in a computationally intensive framework. The ability of these models to predict first-episode psychosis, which is a novelty and one of the contributions of this paper, has been extensively verified through large scale Monte Carlo simulations in the same computationally intensive framework.

Then, the predictive value of cannabis related variables with respect to first-episode psychosis was demonstrated in this work by showing that there is a statistically significant difference between the performance of the predictive models built with and without cannabis variables. We were inspired in this approach, proposed and implemented here, by the Granger causality techniques [13], which are used to demonstrate that some variables have predictive information on other variables in a regression context, as opposed to classification, which is mainly the case in our framework.

Finally, we extended our approach by further investigating how different patterns of cannabis use relate to new cases of psychosis, via association analysis and bayesian techniques such as Apriori and Bayesian Networks, respectively.

The rest of the paper is organized as follows. Section 2 presents our approach to predicting the first-episode psychosis, based on experimenting with various machine learning algorithms and on computational intensive model optimisations. The section includes also the data pre-processing, and investigates the outcomes of the extensive Monte Carlo simulations in order to study the variation of the model performances that may have been affected also by the presence of a high proportion of missing values in the data. In Section 3, we build optimised prediction models without the cannabis attributes to study if there is a statistically significant difference between the predictive power obtained with and without the cannabis attributes. Then, we investigate and discuss further relationships between the cannabis variables and the first-episode psychosis. Finally, the directions for future work and the conclusion are presented in Section 4.

II. Predicting First-Episode Psychosis: a Computationally Intensive Approach

The data used to develop our novel approach to predicting the first-episode psychosis is a part of a case-control study at the inpatient units of the South London and Maudsley NHS Foundation Trust [4]. The clinical data consists of 1106 records, including 489 patients, 370 controls and 247 unlabeled records. Those described as patients were patients of the Trust who at one time presented with first-episode psychosis; controls were recruited from the same local area by a dedicated research team. Each record refers to a participant of the study and has 255 possible attributes, which were divided into four categories. The first category consists of demographic attributes which represent general features such as gender, race, and level of education. Secondly, drug-related attributes contain information on the use of non-cannabis drugs such as tobacco, stimulants and alcohol. The third category is formed of genetic attributes which were removed from the analysis for the purpose of this study. The final category contains cannabis-related attributes such as the duration of use, frequency, cannabis type, etc.

In order to build our approach to predicting the first-episode psychosis, the data required a set of pre-processing transformations, including feature selection, data sampling, data type conversions as needed by training certain types of models, and missing value imputation.

The prediction modelling consisted of considering various machine learning techniques which are suitable for this classification problem and the dataset, including k-Nearest Neighbours, Support Vector Machine with different kernels, Decision Trees, Bagged Trees, Boosted Classification Trees, eXtreme Gradient Boosting and Random Forests [6], [14]. The models were evaluated based on accuracy, area under curve, precision, sensitivity, specificity, and Cohen's kappa statistic. All experiments, including model training and optimisation based on repeated cross validations, and extended Monte Carlo simulations based on split validations to investigate the stability of the performances of the models, were conducted in a computationally intensive framework, using the packages R, Rapidminer, Weka and Apache Spark, by performing a parallel processing on a Data Science cluster of 11 servers based on Xeon processors and 832GB of RAM.

A. Data Pre-Processing

Data pre-processing was performed before modelling in order to rationalise the complexity of the data and prepare the data for use. The pre-processing consisted of the stages of rationalisation and refinement.

1) Rationalisation: The work of this stage sought to perform a high-level simplification of the dataset, and included several steps. First, records that were missing critical data

Table I
CANNABIS USAGE ATTRIBUTES

Attribute	Description
lifetime_cannabis_user	Ever used cannabis: yes or no
age_first_cannabis	Age when first used cannabis: 7 to 50
age_first_cannabis_under15	Age less than 15 when first used cannabis: yes, no or never used
age_first_cannabis_under14	Age less than 14 when first used cannabis: yes, no or never used
current_cannabis_user	Current cannabis user: yes or no
cannabis_fqcy	Pattern of cannabis use: never used, only at weekends, or daily
cannabis_measure	Cannabis usage measure: none, hash less than once per week, or hash at weekends, hash daily, skunk less than once per week, or skunk at weekends, skunk daily
cannabis_type	Cannabis type: never used, hash, or skunk
duration	Cannabis use duration: 0 to 41 (months)

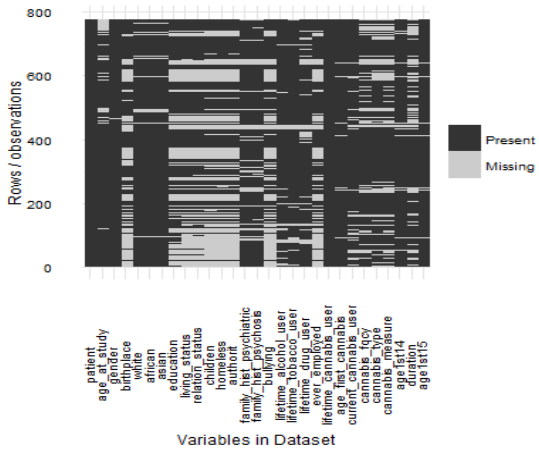

Figure 1. Summary of the ratio of missing values for each attribute

were removed from the dataset. This included records with missing labels, as well as records for which all cannabis-related variables were missing. Secondly, certain variables were removed from the data. This primarily involved variables that were deemed to be irrelevant to the study (such as those related to individual IDs of the study participants), and also included variables which were outside the scope of the current study (for example, certain gene-related variables). In addition, any numeric predictors that had zero or near-zero variance were dropped. Lastly, we sought to make the encoding of missing values consistent across the data set.

2) Refinement: This stage requires several steps: first, the variables were re-labelled to provide more intuitive descriptions of the data contained within. Then, since in multiple situations some variables had a similar meaning as other variables, yet there were often missing values for some records in some of these variables, a process of imputation was used to effectively combine the information from related variables into one. Finally, any attribute that contained more than 50% missing values was dropped from the study. We then removed any record for which more than 70% of the remaining attributes contained missing values. The resulting dataset, after the transformations above, contained 777 records and 29 attributes. The records are divided into: 446 patients and 331 controls. A summary of some of these fields - specifically the ones that relate to cannabis use such as type, age first use, and duration - can be seen in Table I.

B. Missing Values and Imputation

Although the dataset was pre-processed and attributes with more than 50% missing values have been removed from the dataset, it still contains a large number of missing values. Of the 777 records, only 22.5% are complete records. This volume of missing information makes modelling more challenging, but often is the reality of medical and social research. In this study, we used a superset of the variables that were explored in [4], in order to examine the efficacy of machine learning to deal with a significant number of missing values present on the whole collected dataset. A plot of the proportion of missing values is shown in Figure 1. Only two variables - the output attribute *patient* and the input attribute *lifetime_cannabis_user* - are populated completely. On the other extreme, the attribute *children* is missing values in 48% of the records.

Missing values can exist in medical data sets for many reasons such as some participant may be unable to fully complete a survey, or may want to abstain from answering certain questions, or do not attend follow-up appointments, etc. Alternatively, researchers may decide to add or remove certain attributes from the data collection process over time. Missing values mostly need to be imputed before applying pattern discovery techniques. However, the predictive power of the data may depend significantly on the way missing values are treated. While some machine learning algorithms, such as decision trees [6], have the capability to handle missing data outright, most machine learning algorithms do not. Usually, in medical research applications missing values are imputed using a supervised learning technique such as k-Nearest Neighbour. These imputation techniques do not have theoretical formulations but have been much implemented in practice [15].

In this work, we opt for the *tree bagging imputation* from the *caret* package [14] to impute the missing values in the training data sets which are generated in the cross validations, or in the repeated experiments of the Monte Carlo simulations. This imputation process is thus repetitive, becomes a part of each model training and is therefore evaluated as part of each model's performance metrics. We should note also those imputations performed with k-Nearest Neighbour led finally to slightly weaker predictive results than those based on tree bagging imputations, although the latter are more computationally costly as based on an ensemble technique. This of course matters in a computationally intensive framework comprising an intensive model

Table II
SUMMARY OF PARAMETERS TUNED FOR EACH MODEL

Model	Optimised Parameter(s)
C5 Decision Tree	Iter = 70 , Model = rules Winnow = False
Boosted Trees	Iter = 100, MaxDepth= 5, Nu = 0.1
eXtreme Gradient Boosting	Nrounds = 50, MaxDepth = 3 Eta= 0.3 , Gamma= 0 , Colsample= 0.8, MinChild =1
Bagged CART	*None*
Random Forests	Mtry = 30
SVM (Linear)	Cost = 16
SVM(Radial)	Cost = 16384 , Gamma = 3.05e-05
SVM (Poly)	Cost = 64, Degree = 1, Scale= 0.1
KNN	K = 5

optimisation and extensive Monte Carlo simulations. As such the use of adequate computing power is the solution we benefited of to handle also this aspect.

As a final transformation of the data, since some prediction modelling algorithms, such as Support Vector Machines, work only with numeric data, we transformed the input nominal variables into dummy variables, obtaining a dataset of 91 variables.

C. Training and Optimising Predictive Models

For the purpose of developing optimised predictive models for the first-episode psychosis, the values of the parameters for each of the algorithms have been controlled by suitably chosen grids. Predictive models have been fitted, in a 10-cross validation procedure, on each training set after tree bagging imputations of missing values on the same training set, and have been tested on each test set. The best performance models with their parameters have been selected for subsequent comparison. For instance, Support Vector Machine algorithm has been tuned for different kernels and values for the cost and gamma parameters. The best model was obtained on the Radial (RBF) kernel, with the cost 16384, and gamma $3.052e{-}05$.

A summary of the models with the chosen optimised parameters is shown in Table II. The key performance measure was the *accuracy* (the rate of accurate classification) and we monitored also Kohen's *kappa* statistic (the agreement between actual values and predictions, adjusted for what could be expected from pure chance). Values of initial estimates of accuracy and kappa for all models can be seen in Table III.

Based on these results, the models that were selected for further analysis were Boosted Classification Trees (AdaBoost), Random Forests and Support Vector Machine with a Radial kernel.

D. Monte Carlo Simulations

Due to expected potential variations of the predictive models' performance, depending on the datasets for training and testing, but in particular due to the uncertainties introduced by the missing values in the data, we conducted extensive Monte Carlo simulations to study these variations, and thus the stability of the models. In particular the simulations for each of the three algorithms, Boosted Classification Trees, Random Forests, and Support Vector Machine with a Radial kernel, consisted of 10,000 stratified split validations with a 3/4 and 1/4 split for training and testing sets, respectively. On each training set a tree bagging imputation was performed prior to fitting a model with its corresponding optimal parameters. The models' performances consisting of accuracy, precision, sensitivity, specificity, area under the curve, and kappa were estimated on the test set in each iteration. The aggregation of all iterations formed various distributions of the above performance measures.

Table III
INITIAL ESTIMATION OF MODEL ACCURACY & KAPPA

Model	Accuracy	Kappa
SVM (Radial)	0.7824	0.5519
AdaBoost	0.7850	0.5586
Random Forests	0.7798	0.5465
xgbTree	0.7786	0.5432
C5 Decision Tree	0.7682	0.5211
Bagged CART	0.7540	0.4950
SVM (Linear)	0.7619	0.5078
SVM (Poly)	0.7657	0.5160
KNN	0.6923	0.5074

Figure 2 shows histogram plots of the Monte Carlo simulations for highest-performing models based on Boosted Classification Trees. These models achieved a mean accuracy of 79.0% (95% CI [.73, .84]) and a mean kappa of .56 (95% CI [.45, .66]). The latter shows significant predictive information of the input attributes over the first-episode psychosis.

We remark a good predictive power and stability of these models, based on an acceptable level of variation of their performance measures evaluated across extensive Monte Carlo experiments.

III. CANNABIS ATTRIBUTES' PREDICTIVE INFORMATION OVER FIRST-EPISODE PSYCHOSIS

After performing Monte Carlo simulations, the best performing models have been further analysed in order to better understand the predictive power of the cannabis-related attributes over the first-episode psychosis. Moreover we have investigated the link between cannabis-related attributes and the first-episode psychosis via association analysis and bayesian inference based techniques.

A. Predicting First-Episode Psychosis Without Cannabis Attributes

We re-fit best-performing models with the three chosen algorithms (Boosted Classification Trees, Random Forests and Support Vector Machine with a Radial kernel) but this time with the cannabis-related attributes, represented

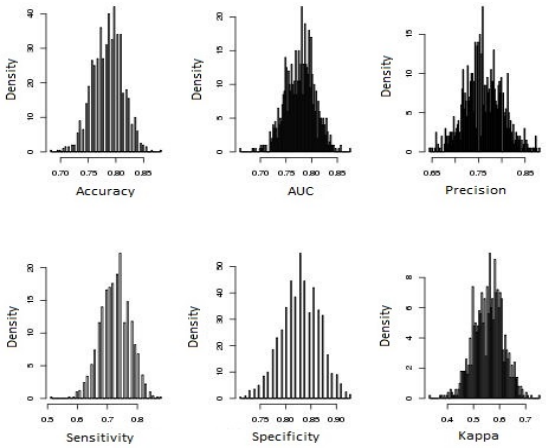

Figure 2. Monte Carlo simulation for Boosted Classification Trees

in Table I, removed from the dataset. The performances obtained with and without the cannabis-related attributes are compared using Student's t-test. That is, the predictive value of cannabis-related attributes with respect to first-episode psychosis is demonstrated by showing that there is a statistically significant difference between the performances of the predictive models built with and without the cannabis variables. We were inspired in this approach we propose here, by the Granger causality techniques [13], which are used to demonstrate that some variables have predictive information on other variables in a regression context (as opposed to classification, in this case).

Our analysis shows that the accuracy of all models decreased around 7% if the cannabis-related attributes are removed from the process of building the predictive models. If we compare, for instance, the accuracies of the best two Random Forests models obtained on the data sets with and without the cannabis attributes, the p-value obtained for the one-tailed t-test was 0.0003. We conclude that the model *with* cannabis attributes has higher predictive accuracy. In other words, the additional cannabis variables jointly account for predictive information over the first-episode psychosis.

B. Cannabis Use and First-Episode Psychosis Associations

To further explore the link between the cannabis attributes and the first-episode psychosis, we look into detecting patterns in data with association analysis and bayesian inference techniques such as Apriori [16] and Bayesian Networks [17], respectively.

A repetitive fine tuning of Apriori led to the detection of the top 6 rules represented in Figure 3. The quality of these rules is expressed by their confidence estimates, and by 95% confidence intervals for these estimates. The rules represent patterns in the general local population in the mentioned area, since the data sample is representative of this area's

1- Cannabis user=Yes & Cannabis=Daily & Cannabis type=Skunk → Patient
(conf=.85, 95%CI[.773,.899]
2- Cannabis user=Yes & age1st14=Yes & Cannabis type=Skunk → Patient
(conf=.81, 95%CI[.723,.87]
3- Cannabis user=Yes & Cannabis type=Skunk → Patient
(conf=.79, 95%CI[.72,.839]
4- Cannabis user=Yes & Cannabis=Daily → Patient
(conf=.74, 95%CI[.67,.793]
5- Cannabis user=Yes & Cannabis=Daily & ageFirstCannabis=15 → Patient
(conf=.73, 95%CI[.643,.805]
6- Cannabis user=Yes & Duration =above 6 → Patient
(conf=.71, 95%CI[.634,.778]

Figure 3. Top association rules

population. The first rule states that if a participant were a cannabis user who consumes skunk daily, then there is an 85% likelihood that this participant is a first-episode psychosis patient. This rule shows evidence of a strong association between using skunk daily, and the first-episode psychosis. If this type of cannabis is used daily or less often, then the likelihood of the first-episode psychosis decreases from 85% to 79%, as expressed by rules 1 and 3 together. For a general type of cannabis which is used daily, the likelihood of a user to be a first-episode psychosis patient decreases from 85% to 74%, as expressed by rules 1 and 4 together. Rule number 2 supports findings from [2], [4] by associating the age of the first use of high potency cannabis with the psychosis onset. Rule number 5, having a 73% confidence, is consistent with findings from [18] regarding the onset of psychosis among cannabis users in relation to a cannabis consumption starting at the age of 15 and younger. Finally, rule number 6 expresses the finding that if a participant were a cannabis user who used cannabis for at least 6 months, then there is a 71% likelihood that this participant is a first-episode psychosis patient.

Recently, Bayesian Networks have been used in Psychiatry as a decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment [17]. We have applied this machine learning technique to further detect the interaction between the first-episode psychosis and the cannabis attributes. As such only cannabis-related attributes were used as predictors in fitting the Bayesian Network model, whose DAG is depicted in Figure 4. The model details suggest that *duration* and *cannabis_type* are among the most predictive attributes. The Bayesian Network probability distributions (not included here due to lack of space) show that subjects who started using cannabis by the age of 15, and consumed cannabis daily for more than 6 months, are twice more likely to be patients rather than controls. On the other hand, subjects who started using cannabis by the age of 15 and consumed cannabis only at the weekend for more than 6 months, increase their chance by 1.5 times to be patients rather than controls. In addition, the model confirms that subjects who used skunk daily are twice more likely to be patients rather than controls. These findings support the idea that cannabis use could lead to the onset of the first-episode psychosis.

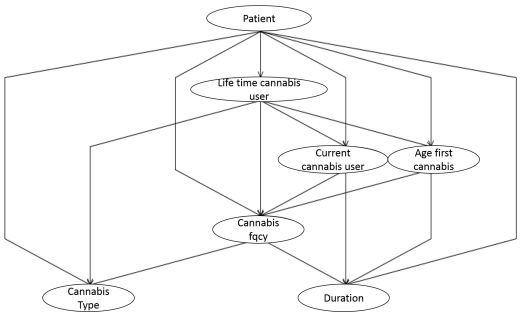

Figure 4. Bayesian Network for cannabis variables

IV. Conclusion and Directions of Further Work

The aim of this work has been to propose a novel symbiotic machine learning and statistical approach to pattern detection and to developing predictive models for the onset of the first-episode psychosis. The predictive models we developed in this novel approach have been optimised in a computationally intensive framework. They exhibited a good predictive power and stability based on an acceptable level of variation of their performance measures evaluated across extensive experiments encapsulated in a series of large-scale Monte Carlo simulations.

A significant proportion of the models performance variation may be explained by the uncertainties present in the data, represented by the high proportion of missing values. As such it would be interesting to further investigate how this prediction performances variation evolves with the thresholds for missing data used in the selection of the attributes and records

Another direction of research that will be explored next consists of investigating the impact of including genotype data in the study, and redefining the predictive modelling approach by taking into account the particularities of the newly introduced data, such as the high dimensionality.

Finally, we are currently investigating into the first-episode psychosis predictiveness enhancements by considering Artificial Neural Network and Deep Learning approaches.

References

[1] T. Moore, S. Zammit, A. Lingford-Hughes, et al., "Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review", The Lancet, vol. 370, no. 9584, pp. 319-328, 2007.

[2] M. Di Forti, C. Morgan, P. Dazzan, et al., "High-potency cannabis and the risk of psychosis", The British Journal of Psychiatry, vol. 195, no. 6, pp. 488-491, 2009.

[3] S. Dragt, D. Nieman, F. Schultze-Lutter, et al., "Cannabis use and age at onset of symptoms in subjects at clinical high risk for psychosis", Acta Psychiatrica Scandinavica, vol. 125, no. 1, pp. 45-53, 2011.

[4] M. Di Forti, A. Marconi, E. Carra, et al., "Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study", The Lancet Psychiatry, vol. 2, no. 3, pp. 233-238, 2015.

[5] R. Radhakrishnan, S. Wilkinson and D. DSouza, "Gone to Pot: A Review of the Association between Cannabis and Psychosis", Frontiers in Psychiatry, vol. 5, 2014.

[6] M. Kuhn and K. Johnson, "Applied Predictive Modeling", Springer, 2013.

[7] H. Zhou, J. Tang and H. Zheng, "Machine Learning for Medical Applications", The Scientific World Journal, vol. 2015, pp. 1-1, 2015.

[8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines", Machine Learning, vol. 46, no. 1-3, pp. 389-422, 2002.

[9] R. Iniesta, D. Stahl and P. McGuffin, "Machine learning, statistical learning and the future of biological research in psychiatry", Psychological Medicine, vol. 46, no. 12, pp. 2455-2465, 2016.

[10] J. Chou, J. Tsai, and T. Liu, "Predictive models for 5-year mortality after breast cancer surgery", Machine Learning and Cybernetics (ICMLC), vol. 1. IEEE, pp.13-16, 2014.

[11] K. Foster, R. Koprowski, et.al, "Machine learning, medical diagnosis, and biomedical engineering research - commentary", BioMedical Engineering OnLine, vol. 13, no. 1, p. 94, 2014.

[12] Kapur, A. Phillips and T. Insel, "Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it?", Molecular Psychiatry, vol. 17, no. 12, pp. 1174-1179, 2012.

[13] M. Ding, Y. Chen and S.L. Bressler. "Granger causality: Basic theory and application to neuroscience", Handbook of Time Series Analysis, pages 451474. Wiley-VCH Verlage, 2006.

[14] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt et al., "caret: Classification and Regression Training", R package version 6.0-64, 2016.

[15] G. E. Batista and M. C. Monard, "An analysis of four missing data treatment methods for supervised learning", Applied Artificial Intelligence, vol. 17, no. 5-6, pp. 519-533, 2003.

[16] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases", In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487–499, Santiago, Chile, September 1994.

[17] F. Seixas, B. Zadrozny, J. Laks et al., "A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimers disease and mild cognitive impairment", Computers in Biology and Medicine, vol. 51, pp. 140–158, 2014.

[18] M. Di Forti, H. Sallis, F. Allegri et al., "Daily Use, Especially of High-Potency Cannabis, Drives the Earlier Onset of Psychosis in Cannabis Users", Schizophrenia Bulletin, vol. 40, no. 6, pp. 1509–1517, 2013.